Genomewide prediction of genotypic values and genetic variances within 969 maize biparental populations

A THESIS SUBMITTED TO THE FACULTY OF

UNIVERSITY OF MINNESOTA

BY

Lian Lian

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Rex Bernardo

September 2014

Acknowledgements

I am very thankful to my advisor Dr. Rex Bernardo, who gave me an opportunity to work on my PhD projects. Dr. Bernardo is a great mentor and he ignited my interest in plant breeding/quantitative genetics. He gave me a lot of freedom to experiment out different ideas and in the meantime provided me with supportive and insightful guidance. Additionally I would thank my thesis committee: Dr. James Anderson, Dr. Kevin Smith, Dr. Aaron Rendahl, and Dr. Shengqiang Zhong for all their involvement in my degree completion.

I thank Monsanto for proving financial support and the data for this study. Without their support, this study could not have been done and published.

I am also thankful for my family and friends. The journey through graduate school was beautiful and cherished with their support and company.

Abstract

In plant breeding, selecting within biparental crosses and selecting parents to make new crosses are both important. My first study investigated the accuracy of genomewide selection (r_{MG}) within 969 biparental maize populations (*Zea mays* L.). My objectives were to determine: (i) the mean and variability of r_{MG}, (ii) if r_{MG} can be predicted, and (iii) how training population size (N), heritability (h^2), and number of markers (N_M) affect r_{MG}. I modified an equation for expected r_{MG} [E(r_{MG})] to account for linkage disequilibrium (r^2) between markers and quantitative trait loci. Across the 969 populations, the mean and range (in parentheses) of observed r_{MG} was 0.45 (–0.59, 1.03) for yield, 0.59 (–0.34, 0.96) for moisture, and 0.55 (–0.24, 1.10) for test weight. The observed r_{MG} values were centered around E(r_{MG}) when r^2 was accounted for, but had a large spread around E(r_{MG}). The $r^2(Nh^2)^{½}$ had the strongest association with the observed r_{MG}. In the second study, my objective was to determine whether related populations could be used to predict the genetic variance (V_G) of a segregating population from two parents (A and B). For each of 85 A/B populations, 2–23 A/* and B/* populations were used as training populations, where * denotes a random parent. In the genomewide selection model, the testcross V_G in A/B was predicted as the variance among the predicted genotypic values of progeny from the simulated A/B population. In the mean variance model, V_G was estimated as the mean of V_G in A/* and B/* populations. The correlations between observed and predicted V_G were not significant ($P = 0.05$) for the

genomewide selection model but were significant for the mean variance model (0.26 for yield, 0.46 for moisture, and 0.50 for test weight). The V_G of A/B population could therefore be predicted as the mean of V_G in A/* and B/* populations. Overall, the results indicated that genomewide selection can identify the best individuals within a cross, but it cannot reliably predict which parents would lead to the largest genetic variance.

Table of Contents

List of Tables

Chapter 1

Chapter 2

List of Figures

Chapter 1

Chapter 1: Genomewide prediction accuracy within 969 maize biparental populations

In genomewide selection, the expected correlation between predicted and true genotypic values (r_{MG}) has been previously derived as a function of the training population size (N), heritability (h^2), and effective number of chromosome segments (M_e) affecting the trait. Our objectives were to determine: (i) the mean and variability of r_{MG} in 969 biparental maize (*Zea mays* L.) breeding populations for seven traits, (ii) if r_{MG} can be predicted in advance, and (iii) how N, h^2, and number of markers (N_M) affect r_{MG}. We modified a previous equation for expected r_{MG} to account for linkage disequilibrium (r^2) between a marker and a quantitative trait locus (QTL). Across the 969 populations, the mean and range (in parentheses) of observed r_{MG} was 0.45 (–0.59, 1.03) for grain yield, 0.59 (–0.34, 0.96) for moisture, 0.55 (–0.24, 1.10) for test weight, 0.49 (–0.22, 1.04) for stalk lodging, 0.41 (–0.30, 0.93) for root lodging, 0.47 (–0.45, 0.97) for plant height, and 0.42 (–0.43, 0.94) for ear height. The observed r_{MG} values were centered around the expected r_{MG} when r^2 was accounted for, but the observed r_{MG} had a large spread around the expected r_{MG}. The $r^2(Nh^2)^{½}$ had the strongest association with observed r_{MG}. When $r^2(Nh^2)^{½}$ exceeded 8, the proportion of r_{MG} equal to or larger than 0.50 reached 90% among all the population–trait combinations. We conclude it is difficult to predict r_{MG} in advance, but that rules of thumb based on $r^2(Nh^2)^{½}$ can help achieve a high r_{MG}.

Introduction

Maize breeding typically involves crossing two inbred parents, developing inbred progeny from the biparental cross, and selecting the best new inbreds on the basis of testcross performance (Hallauer, 1990). Traditionally, selection for complex traits has been done by field evaluation of testcrosses in multiple environments. Genomewide selection (or genomic selection) (Meuwissen et al., 2001) allows the prediction of genotypic values of individuals for complex traits on the basis of marker information. In genomewide selection, marker effects are estimated from a training population that has been genotyped and phenotyped. Genotypic values of individuals in a test population that has been genotyped but not phenotyped, are then predicted from the marker effects estimated from the training population.

Different breeding schemes for genomewide selection have been proposed and studied in maize and in other species (Bernardo and Yu, 2007; Bernardo, 2009; Heffner et al., 2010; Riedelsheimer et al., 2012, 2013; Windhausen et al., 2012). Regardless of the breeding scheme, the r_{MG} must be high enough for genomewide selection to be time and cost effective. The expected prediction accuracy [$E(r_{MG})$] has been previously derived as a function of N, h^2 and M_e affecting the trait (Daetwyler et al., 2008):

$$E(r_{MG}) = [Nh^2/(Nh^2 + M_e)]^{1/2} \qquad [1]$$

The M_e pertains to the idealized concept of having independent chromosome segments, with each segment containing a QTL-marker pair and with all the QTL having additive, equal effects (Daetwyler et al., 2008; Goddard, 2009). When the genome is saturated by

markers, M_e is generally calculated on the basis of the effective population size and the genome size (Goddard, 2009; Meuwissen and Goddard, 2010; Lorenz, 2013). The M_e can also be calculated by the eigenvalues of the marker correlation matrix, according to the same approach for calculating the effective number of independent tests in association mapping when the markers are highly correlated (Li and Ji, 2005).

Before they commit time and resources to genomewide selection, breeders want to know whether or not r_{MG} will be high enough in a population. Previous empirical studies on the correspondence between observed and predicted r_{MG} have been few and were limited by the number of populations studied. In five maize biparental crosses, the observed r_{MG} agreed well with the r_{MG} expected from Eq. [1] (Riedelsheimer et al., 2013). In four crosses in maize, barley (*Hordeum vulgare* L.) and *Arabidopsis thaliana* (L.) Heynh, the observed r_{MG} generally agreed with $E(r_{MG})$ but uncertainty in M_e made the comparisons difficult (Combs and Bernardo, 2013a).

The current study utilizes phenotypic and marker data from 969 biparental maize populations in the Monsanto breeding program. The data in this study therefore represent the genetic backgrounds, traits, h^2, population sizes, and extent of testing in a commercial breeding program. As such, the data can provide a realistic indication of the r_{MG} for different traits in maize and of the correspondence between the observed and expected r_{MG} when genomewide selection is routinely practiced on a wide scale. To avoid any confounding effects of a difference in the genetic constitution of the training population and test population (Daetwyler et al., 2008), we analyzed each of the 969 biparental

crosses individually. Our objectives were to determine: (i) the mean and variability of r_{MG} in maize biparental populations, (ii) if r_{MG} can be reliably predicted in advance, and (iii) the how r_{MG} is affected by traits, h^2, N, and N_M in biparental populations.

Materials and Methods

Phenotypic and marker data

The 969 maize populations comprised two maize heterotic groups, with 485 biparental crosses in Group 1 and 484 crosses in Group 2. The number of lines in each cross ranged from 35 to 356 and had a mean of 156. The lines in each cross were derived from the following generations (number of crosses in parentheses): F_2 (707), BC_1 (186), BC_1F_2 (47), doubled haploid (DH) from F_1 (17), and DH from F_2 (12). For the F_2, BC_1, and BC_1F_2 populations, plants grown from 10 to 15 selfed seeds derived from each individual plant were testcrossed to an inbred tester from the opposite heterotic group. For DH populations, each DH line was crossed to an inbred tester from the opposite heterotic group.

The populations were evaluated for the following traits: grain yield (Mg ha^{-1}), moisture (g H_20 kg^{-1}), test weight (kg hl^{-1}), stalk lodging (%), root lodging (%), plant height (cm), and ear height (cm). Each experiment was conducted in 2 to 15 locations (usually 6–8 locations) during a single year (2000–2008). Phenotypic data were available as the testcross mean of each line at each location. Not all traits were measured in all locations and in all populations (Table 1). To help ensure that the lodging traits were

adequately expressed within a location, only those locations with a mean stalk lodging of at least 5% or a mean root lodging of at least 5% were retained for these two traits, and a population was retained for these two traits only if data were available for more than one location.

The lines within each cross were genotyped with 31 to 119 (mean of 70) SNP markers that were polymorphic between the two parents and that were a subset of the 2911 SNP markers used for genotyping all the parents. When the set of markers (31 to 119 SNP loci) used for each population did not have data for some lines, the missing marker data were imputed with fastPHASE (Scheet and Stephens, 2006); imputation with the full set of 2911 parental markers was not done. A segregating SNP locus was disregarded if the parents were monomorphic, if the minor allele frequency was < 0.10, or if more than half of the data points were missing. Populations with < 30 markers were removed from analysis.

Heritability

Within each cross, testcross genetic (V_G) and nongenetic (V_R) variance components were calculated for each trait by restricted maximum likelihood via the "lmer" function in the "lme4" package (Bates et al., 2013). Because the data were entry means within each location, the genotype by environment interaction variance and within-location error variance were confounded in V_R. A likelihood ratio test was used to test the significance of the estimates of testcross genetic variance. The *p*-value from the likelihood ratio test was divided by 2.0 to approximate an F-test of the null hypothesis

(Holland et al., 2003). Genetic variance estimates with a p-value > 0.05 were considered not significant and the corresponding population–trait combination was discarded. The data for each biparental cross were not completely balanced because some individuals were not evaluated at a few of the locations in each experiment. As such, the h^2 was estimated on an ad hoc basis as $h^2 = V_G/(V_G + V_R/l)$, where l was the harmonic mean of the number of locations (Holland et al., 2003).

Observed r_{MG}

Genomewide marker effects were obtained by ridge-regression best linear unbiased prediction (RR-BLUP) as described by Meuwissen et al. (2001) and by Bernardo and Yu (2007). Suppose the phenotypic data comprised N_T records, each record being the mean performance of an individual in one of the N_L locations. The linear model was as $\mathbf{y} = \mathbf{Xb} + \mathbf{Zm} + \mathbf{e}$, where $\mathbf{y}$ was an $N_T \times 1$ vector of phenotypic records, $\mathbf{b}$ was an $N_L \times 1$ vector of fixed effects of locations, $\mathbf{m}$ was an $N_M \times 1$ vector of random effects of the markers, $\mathbf{e}$ was an $N_T \times 1$ vector of residuals, $\mathbf{X}$ was an $N_T \times N_L$ incidence matrix that related $\mathbf{y}$ to $\mathbf{b}$, and $\mathbf{Z}$ was in $N_T \times N_M$ incidence matrix (with values of 1 and –1 for each of the two homozygotes and 0 for the heterozygote) that related $\mathbf{y}$ to $\mathbf{m}$. The variance of marker effects was V_G/N_M (Meuwissen et al., 2001).

The r_{MP} for each trait within each population was estimated by a delete-one method (Kohavi, 1995). Suppose a population had N_P = 100 lines that had been phenotyped. The first line was assumed untested and its performance was predicted from genomewide marker effects estimated from RR-BLUP analysis of lines 2 to 100. The

second line was then assumed untested and its performance was predicted from RR-BLUP analysis of the remaining $N = N_P - 1$ lines. In the end, the correlation between the predicted genotypic value and the mean phenotypic value of the N_P lines was calculated and was denoted by r_{MP}. The value of r_{MG} was calculated as r_{MP} divided by the square root of h^2 (Dekkers, 2007). For each r_{MP}, we calculated the test statistic $T = [r_{MP}(N_P - 2)^{½}]/(1 - r_{MP}^2)]$, which follows a t_{N-2} distribution (Bobko, 2001). Significance tests for r_{MP} (via T) were done through a t-test. Given that significance tests for genetic variances had been previously done, we assumed for simplicity that r_{MG} was significantly different from zero if r_{MP} was significantly different from zero.

Expected r_{MG} with incomplete linkage disequilibrium

Eq. [1] was derived with the following four assumptions (Daetwyler et al., 2008). First, the marker effects were derived from simple linear regression rather than from RR-BLUP, which we used in this study. Second, each marker–QTL pair was assumed independent of other marker-QTL pairs. Third, the different marker–QTL pairs (which were to be accounted for by M_e) were assumed to have equal variances. Fourth, each marker–QTL pair was assumed in complete linkage disequilibrium. We modified Eq. [1] to retain the first three assumptions by Daetwyler et al. (2008) but to relax the last assumption. By accounting for incomplete linkage disequilibrium between a marker and QTL, we obtained the following $E(r_{MG})$ with $r^2 < 1$ (Appendix 1 in Supplemental Material):

$$E(r_{MG\_r2}) = r^2[Nh^2/(r^2Nh^2 + M_e)]^{1/2} \qquad [2]$$

In Eq. [2], we used $r^2 = r^2_{MM/2}$, the latter being the mean squared correlation between a marker and QTL when the QTL is assumed to be at the midpoint of the two markers. The midpoint, in turn, is the mean QTL position assuming a uniform distribution between two markers. The r^2_{MM} can be calculated as the square root of the observed squared correlation between two marker genotypes. We found that $r^2_{MM/2} = \sqrt[2]{r^2_{MM}}$ for BC_1, F_2, and DH populations but not for BC_1F_2 populations (Appendix 2 in Supplemental Material). The $E(r_{MG\_r2})$ values were therefore not calculated for the 47 BC_1F_2 populations. For all the following analyses, we used $r^2 = r^2_{MM/2}$. Although the assumption of complete linkage between a marker and a QTL was relaxed, Eq. [2] assumed that the distance between a QTL and a marker is constant for all marker–QTL pairs.

For each population, the M_e values were calculated by eigenvalues of the linkage disequilibrium matrix among the SNP markers in each population (Li and Ji, 2005). This matrix was obtained by computing all pairwise squared correlations for markers on each individual chromosome. Eigenvalue decomposition was done with the "eigen" function in R (R Development Core Team, 2012). Given the known values of N and the estimated values of h^2, r^2, and M_e, the expected r_{MG} was calculated according to both Eq. [1] and [2].

Because of the uncertainty in the proper form of M_e, we also back-calculated M_e by equating the observed r_{MG} with the expected r_{MG} from Eq. [1] and [2] and solving for M_e. This procedure was performed for all population–trait combinations (excluding BC_1F_2) where the observed r_{MG} was between 0.1 and 1.

Factors affecting r_{MG}

We evaluated the relative importance of the following factors in terms of the variance in r_{MG} explained by each factor: trait, h^2, training population size (N being equal to $N_P - 1$), N_M, heterotic group, and generation type. Given their prominence in Eq. [1] or [2], we evaluated the variance explained by $(Nh^2)^{1/2}$ and $r^2(Nh^2)^{1/2}$. For each trait, we also obtained the correlation coefficient between the observed r_{MG} and the following: $(Nh^2)^{1/2}$, $r^2(Nh^2)^{1/2}$, $E(r_{MG})$ from Eq. [1], and $E(r_{MG\_r2})$ from Eq. [2].

To further investigate the effect of N on r_{MG}, 1533 population–trait combinations with rounded (to the nearest tens digit) values of $N_P = 180$ were selected. For each population–trait combination, a random subset of $N_P = 61, 91, 121$, and 151 lines were obtained as a new population and the r_{MP} within the new population was estimated by a delete-one method as described before, so that N was 60, 90, 120, and 150. For each population–trait combination, only one random sample for the new population was obtained because the results were to be averaged over 1533 populations. The value of h used for calculating r_{MG} (Dekkers, 2007) was that obtained with the largest N_P.

Similarly, to investigate the effect of N_M on r_{MG}, 45 population–trait combinations whose N_M values were larger than 100 were selected. Markers were thinned to 2/3 of their original density by keeping only the first two markers out of every three consecutive markers. Likewise, markers were thinned to 1/3 of their original density by keeping only the first marker out of every three consecutive markers. The r_{MP} within the new population was estimated by a delete-one method as described before.

The effect of h^2 on r_{MG} was further studied for the population–trait combinations with rounded $N_P = 180$. Heritability values were subdivided into several intervals and the mean r_{MG} in each interval was calculated. For each trait, if the number of populations in an interval was □ 20, the mean r_{MG} values for that trait were not calculated for that interval and the data were not shown for that trait in the interval. Because we only had a limited number of observations that met the above criteria for stalk lodging and root lodging, this analysis was not done for these two traits.

Likewise, the effect of $(Nh^2)^{1/2}$ and $r^2(Nh^2)^{1/2}$ on r_{MG} was examined by subdividing $(Nh^2)^{1/2}$ and $r^2(Nh^2)^{1/2}$ into different intervals. For each trait, if the number of populations in an interval was □ 20, the mean r_{MG} values for that trait were not calculated for that interval and the data were not shown for that trait in that interval. This analysis was not done for stalk lodging and root lodging and not for BC_1F_2 populations.

To assess their distribution, r_{MG} values were divided into different intervals and the proportion of r_{MG} in each interval was plotted against different traits for all population–trait combinations and plotted against different $(Nh^2)^{1/2}$, different $r^2(Nh^2)^{1/2}$, and different values of $E(r_{MG\_r2})$ for population–trait combinations excluding BC_1F_2.

Results and Discussion

Mean and Variability of Observed r_{MG}

Out of 3371 population–trait combinations, 2919 (87%) had an r_{MG} that was significantly different from zero ($P = 0.05$). The mean r_{MG} was 0.52 and the individual

r_{MG} values ranged from –0.59 to 1.10 (Table 1). The mean r_{MP} was 0.37 and the individual r_{MP} values ranged from –0.34 to 0.89 (Table 1). The r_{MG} was estimated indirectly as r_{MP}/h (Dekkers, 2007), and the r_{MG} values that exceeded 1.0 were due to sampling variation in both r_{MP} and h^2. Only 11 of the 3371 population–trait combinations (0.3%) had an r_{MG} that was negative and significantly different from zero ($P = 0.05$).

The mean, range, and standard deviation of r_{MG} values differed among the seven traits. The mean r_{MG} was highest for moisture and lowest for root lodging (Table 1). Conversely, the standard deviation of r_{MG} was smallest for moisture and largest for root lodging. For grain yield, which was the most important trait (Bernardo, 1991), the mean r_{MG} (range in parentheses) across 840 populations was 0.45 (–0.59, 1.03) and around 40% of the populations had an $r_{MG} \geq 0.50$ (Fig. 1a). The middle 50% of r_{MG} values for grain yield ranged from 0.32 to 0.59. For moisture, which was the second most important trait (Bernardo, 1991), the mean r_{MG} (range in parentheses) across 943 populations was 0.59 (–0.34, 0.96) and around 80% of the populations had an $r_{MG} \geq 0.50$. The middle 50% of r_{MG} values for moisture ranged from 0.53 to 0.71.

Among the seven traits studied, stalk lodging and root lodging are the two traits that are generally considered by maize breeders as the least consistent in their expression. However, the mean r_{MG} for stalk lodging (0.49) and the mean r_{MG} for root lodging (0.41) were not significantly different from the mean r_{MG} for grain yield (0.45). Overall, the results from this 969-population study indicated that the mean r_{MG} for yield and other agronomic traits in maize biparental crosses is in the 0.40 to 0.60 range.

Expected r_{MG}

There were 3217 population–trait combinations after removing the BC_1F_2 populations, for which r^2 could not be calculated as the square root of the observed squared correlation between two markers. Whereas the mean r_{MG} across all the 3217 population–trait combinations was 0.52, the mean expected r_{MG} according to Eq. [1] [mean E(r_{MG})] was 0.74. Eq. [1] (Daetwyler et al., 2008), which assumes perfect linkage between a marker and QTL, therefore grossly overestimated r_{MG} (by 0.74 – 0.52 = 0.22). The mean expected r_{MG} according to Eq. [2] [mean E(r_{MG_r2})] was 0.56. Eq. [2], which accounts for imperfect linkage between a marker and QTL, still overestimated r_{MG} but the amount of upward bias (0.56 – 0.52 = 0.04) was much less than that with Eq. [1] from Daetwyler et al. (2008).

The mean r_{MG} was closer to mean E(r_{MG_r2}) for some traits than for others. These mean [r_{MG} – E(r_{MG_r2})] deviations were 0.092 for grain yield, 0.004 for moisture, 0.008 for test weight, 0.021 for stalk lodging, 0.102 for root lodging, 0.058 for plant height, and 0.080 for ear height. Across all traits, the spread of r_{MG} about the mean and about E(r_{MG_r2}) was large (Fig. 2).

Pooled across all traits, the correlation between r_{MG} and E(r_{MG_r2}) was 0.40 whereas the correlation between r_{MG} and E(r_{MG}) was 0.33. The correlation between r_{MG} and E(r_{MG_r2}) was significantly different ($P = 0.05$) from the correlation between r_{MG} and E(r_{MG}) for grain yield, moisture and test weight, but was not significantly different for stalk lodging, root lodging, plant height, and ear height (Table 2). The correlation

between r_{MG} and E(r_{MG_r2}) ranged from –0.17 for root lodging to 0.44 for moisture (Table 2). The correlation between r_{MG} and E(r_{MG}) ranged from –0.32 for root lodging to 0.34 for test weight. The correlation between r_{MG} and E(r_{MG_r2}) was lower for grain yield than for moisture and test weight. Likewise, the correlation between r_{MG} and E(r_{MG}) was lower for grain yield than for moisture and test weight.

The usefulness of Eq. [1] and of Eq. [2] for predicting r_{MG} therefore differed among traits. Overall, the results for observed r_{MG} and expected r_{MG} indicated that although the mean r_{MG} across many different populations and traits can be predicted fairly well by Eq. [2], the r_{MG} for any given trait in any given population cannot be predicted reliably with either Eq. [1] from Daetwyler et al. (2008) or with Eq. [2], which we derived.

Eq. [2] had the following parameters: r^2, N, h^2, and M_e. As shown later, the observed r_{MG} was correlated with $r^2(Nh^2)^{½}$, which was the numerator of Eq. [2]. This result suggested that the failure to accurately predict r_{MG} was mainly due to the failure of M_e to adequately mimic the assumptions of equal and additive effects of marker–QTL pairs. We suggest four other reasons for the inability to predict r_{MG}: (i) while we accounted for imperfect linkage disequilibrium between a QTL and a marker in Eq. [2], this imperfect linkage disequilibrium itself led to not all QTL effects being captured and, consequently, to missing h^2 (Manolio et al., 2009; Makowsky et al., 2011) in the model; (ii) the linear additive model did not capture all of the genetic variance due to epistasis not being modeled; (iii) sampling error in the estimates of r_{MP} and h^2, which were used in

estimating r_{MG}, contributed to a lower observed correlation between r_{MG} and $E(r_{MG\_r2})$; and (iv) uncertainty remains regarding the proper method to calculate M_e.

Previous studies have calculated M_e as $2N_eL$, where N_e is the effective population size and L is the size of the genome in Morgans (Daetwyler et al., 2008). In a maize biparental population, in which all individuals are in theory derived from $N_e = 1$ F_1 plant, the M_e is approximately 30 when M_e is calculated as $2N_eL$. An M_e of 30 always overestimated r_{MG} in the populations used in this study.

The M_e values we used, which were calculated according to Li and Ji (2005), ranged from 28 to 119 and had a mean of 59 in 3034 population–trait combinations (BC_1F_2 populations excluded, and r_{MG} between 0.1 and 1). Because N_M was relatively low, these M_e values did not differ much from the actual N_M, which ranged from 31 to 119 and had a mean of 70. When M_e was back-calculated by equating the observed r_{MG} to $E(r_{MG})$ from Eq. [1] and solving for M_e, the estimated values of M_e ranged from 242 to 525. These M_e values back-calculated from Eq. [1] were much larger than the M_e calculated according to the Li and Ji (2005) method, larger than the actual N_M, and larger than the value of 30 obtained as $2N_eL$ (Table 2). When M_e was back-calculated by equating the observed r_{MG} to $E(r_{MG\_r2})$ from Eq. [2], the estimated values of M_e ranged from 29 to 110. These M_e values back-calculated from Eq. 2 were of similar magnitude to the M_e calculated according to the Li and Ji (2005) method as well as to N_M (Table 2).

Furthermore, the back-calculated M_e values differed among the traits (Table 3). Among the three traits with the most data, grain yield had the largest back-calculated M_e

values whereas moisture and test weight had similar back-calculated M_e values. This result was consistent with the perceived complexity of the traits, with grain yield conceivably being controlled by more QTL compared with moisture and test weight. In contrast, M_e calculated according to the Li and Ji (2005) method or as $2N_eL$ leads to the same M_e across all traits. Overall, these results suggest that other methods need to be developed for calculating M_e.

Association between r_{MG} and Different Factor Combinations

Whereas the best way to estimate M_e is unclear, N in Eq. [1] and [2] were known and h^2 and r^2 were estimated in this study with well-established procedures. We found that $r^2(Nh^2)^{½}$ was most strongly associated with r_{MG}. When only the intercept and a specific factor or a combination of factors was fitted, the variance in r_{MG} explained by the regression model was as follows: $r^2(Nh^2)^{½}$, 17.1%; $(Nh^2)^{½}$, 12.9%; trait, 8.4%; h^2, 7.8%; N, 4.4%; generation type, 0.4%; heterotic group, 0.0%; and N_M, 0.0%. The importance of $r^2(Nh^2)^{½}$ in influencing r_{MG} was in accordance with $r^2(Nh^2)^{½}$ being the numerator of Eq. [2]. When $r^2(Nh^2)^{½}$ exceeded 8, more than 90% of the r_{MG} values were ≥ 0.50 (Fig. 1b).

The correlation between r_{MG} and $r^2(Nh^2)^{½}$ was not significantly different (P = 0.05) from the correlation between r_{MG} and E(r_{MG_r2}) for each trait (Table 2). In other words, r_{MG} was as strongly associated with $r^2(Nh^2)^{½}$ as it was with E(r_{MG_r2}). The correlations of the different factors with r_{MG} were higher for grain yield, moisture, and test weight than for stalk lodging, root lodging, plant height, and ear height. Across all the traits, correlation between r_{MG} and $r^2(Nh^2)^{½}$ was 0.41, between r_{MG} and E(r_{MG_r2}) was

0.40, between r_{MG} and $(Nh^2)^{½}$ was 0.36, and between r_{MG} and E(r_{MG}) was 0.33. Overall, these results indicated that E(r_{MG_r2}) and $r^2(Nh^2)^{½}$ were the best predictors of r_{MG}, and that breeders could manipulate $r^2(Nh^2)^{½}$ to get a high r_{MG} (Fig. 1b).

The mean r_{MG} varied among traits when $r^2(Nh^2)^{½}$ or $(Nh^2)^{½}$ was kept constant (Fig. 3b, c). Grain yield tended to have a lower r_{MG} compared with moisture, test weight, and plant height at a fixed $r^2(Nh^2)^{½}$ or $(Nh^2)^{½}$. For grain yield and plant height, the proportion of r_{MG} equal to or larger than 0.50 did not change much beyond $r^2(Nh^2)^{½} = 6$ and plateaued at around 50% for grain yield and 60 to 70% for plant height. The proportion of r_{MG} equal to or larger than 0.50 increased to above 80% for moisture and test weight when $r^2(Nh^2)^{½}$ exceeded 7.

Association between r_{MG} and Individual Factors

As has been found in previous studies (Lorenzana and Bernardo, 2009; Combs and Bernardo, 2013a; Crossa et al., 2013), increases in the individual factors N, h^2, and N_M generally led to increases in r_{MG}, particularly when it was possible to keep other factors constant. There were 1533 population–trait combinations with the rounded number of lines equal to 180. The mean r_{MG} for different training population sizes was 0.41 with $N = 60$, 0.46 with $N = 90$, 0.51 with $N = 120$, 0.54 with $N = 150$, and 0.55 with rounded $N = 180$. These mean r_{MG} values were significantly different from each other ($P = 0.05$), except for r_{MG} with $N = 150$ vs. rounded $N = 180$.

When the rounded N was fixed at 180, r_{MG} increased as h^2 increased for grain yield, moisture, test weight, plant height, and ear height (Fig. 3a). However, the r_{MG} for

yield did not change much beyond $h^2 = 0.40$. In contrast, the r_{MG} for moisture and plant height keep increasing as h^2 increased.

As mentioned above, N_M explained none of the total variance in r_{MG} when other factors were not kept constant. When all other factors were kept constant and the markers were thinned to 2/3 and 1/3 of their original number, the mean r_{MG} was 0.45 when 1/3 of the markers were used ($N_M = 36$), 0.49 when 2/3 of the markers were used ($N_M = 72$), and 0.50 when all of the markers were used ($N_M = 107$). These mean r_{MG} values across different marker densities were not significantly different from each other ($P = 0.05$). In the above subset of populations, the mean r^2 value between adjacent markers was 0.16 when 1/3 of the markers were used, 0.41 when 2/3 of the markers were used, and 0.52 when all of the markers were used.

A minimum r^2 of 0.20 between adjacent markers has been suggested for genomewide selection (Hayes et al., 2009). The mean r^2 values between adjacent markers was 0.46 across the 969 biparental maize populations used in this study. The r^2 values are typically high in biparental populations because large chromosome segments are passed intact from the inbred parents to the progeny (Smith et al., 2008). The high r^2 between adjacent markers in this study was probably due to having only one meiosis in the development of lines from the F_2 or BC_1 populations, which constituted 92% of the 969 biparental crosses. These high r^2 values suggest that with maize biparental crosses, a fairly small number of SNP markers ($N_M = 70–120$) is largely sufficient for genomewide selection. This agrees with a previous study that found that r_{MG} was at or near maximum

when the mean distance between markers was around 25 cM in a DH maize population (Lorenzana and Bernardo, 2009). On the basis of a linkage map of about 1750 cM (Senior et al., 1996) a 25-cM spacing between adjacent markers is equivalent to having 1750/25 = 70 markers. If the linkage map is smaller because of fewer polymorphic markers between the parents, the N_M equivalent to a 25-cM spacing would be even smaller. On the other hand, the N_M needed would be higher for recombinant inbreds, which would have undergone several meiotic events during their development.

Applications in Plant Breeding

Our results for 969 maize populations show that predicting r_{MG} is difficult. The observed r_{MG} values were centered around the expected r_{MG} when recombination between a QTL and marker was accounted for (Eq. [2]), but the spread of the observed r_{MG} around the expected r_{MG} was large. Breeders should also be aware that r_{MG} as well as the ability to predict r_{MG} differ among traits: grain yield tended to have lower r_{MG} and lower predictability of r_{MG} compared with moisture and test weight.

Our results suggest that r_{MG} is best predicted from both $r^2(Nh^2)^{1/2}$ and $E(r_{MG\_r2})$ from Eq. [2]. The correlations with r_{MG} were equal for $r^2(Nh^2)^{1/2}$ and for $E(r_{MG\_r2})$, but the former cannot predict the actual value of r_{MG}. As a rule of thumb, we recommend $r^2(Nh^2)^{1/2}$ to be at least 8. Such a rule of thumb would lead to about 90% of the r_{MG} values exceeding 0.50. When $r^2(Nh^2)^{1/2}$ is between 5 and 6, about 50% or more of the r_{MG} values would exceed 0.50 for most traits. These rules of thumb apply to using a subset of a biparental cross to predict the performance of a remaining, unphenotyped subset of the

same biparental cross or for recurrent genomewide selection within the same biparental cross (Combs and Bernardo, 2013b; Massman et al., 2013). Other rules of thumb need to be developed for other types of training populations (e.g., pooled biparental crosses; Jacobson et al., 2014). Also, a marker density of 70 SNP loci seems sufficient for F_2 lines or DH lines developed from an elite biparental cross.

We offer a final thought on predicting r_{MG}: we are unable to precisely predict r_{MG} for each population in the same way that breeders are unable to precisely predict h^2, which measures the effectiveness of phenotypic selection. While we know how increasing the number of replications and environments increases the entry-mean h^2, breeders do not devote time in trying to predict h^2. Instead, a breeder designs a yield trial on the basis of knowledge of how the traits vary in different environments, selects lines often without regard for h^2 in the trial, accepts that the outcome of selection decisions will be poor if h^2 is low, but is confident that selection progress can be made when averaged across different populations. We believe that, likewise, breeders should use information of how different factors affect r_{MG}, design a genomewide selection experiment accordingly, be prepared that the outcome of genomewide-selection decision will be poor if r_{MG} happens to be low in a particular test population, but be confident that routine application of genomewide selection across a breeding program will, on average, lead to positive gains. The results from the 969 maize biparental populations in this study should serve as a useful guide in the design of genomewide selection programs.

Table 1. Mean and range of entry-mean heritability (h^2) and prediction accuracy (r_{MP} and r_{MG}) for different traits in 3371 population–trait combinations in 969 maize biparental crosses.

Trait	Populations	h^2		r_{MP}		r_{MG}			
		Mean	Range	Mean	Range	Mean	Range	50% quantiles	Standard deviation
Yield	840	0.46	(0.17, 0.92)	0.30	(–0.34, 0.89)	0.45a[†]	(–0.59, 1.03)	(0.32, 0.59)	0.23
Moisture	943	0.66	(0.24, 0.91)	0.48	(–0.18, 0.81)	0.59c	(–0.34, 0.96)	(0.53, 0.71)	0.19
Test weight	894	0.56	(0.18, 0.92)	0.41	(–0.21, 0.78)	0.55b	(–0.24, 1.10)	(0.46, 0.69)	0.20
Stalk lodging	68	0.33	(0.19, 0.67)	0.28	(–0.13, 0.55)	0.49ab	(–0.22, 1.04)	(0.40, 0.64)	0.24
Root lodging	38	0.32	(0.19, 0.53)	0.23	(–0.17, 0.47)	0.40a	(–0.30, 0.93)	(0.21, 0.67)	0.30
Plant height	369	0.39	(0.18, 0.82)	0.29	(–0.27, 0.69)	0.47a	(–0.45, 0.97)	(0.33, 0.62)	0.24
Ear height	219	0.33	(0.17, 0.63)	0.24	(–0.21, 0.59)	0.42a	(–0.43, 0.94)	(0.27, 0.61)	0.25

[†] r_{MG} values followed by the same letter were not significantly different according to a Tukey HSD test ($P = 0.05$).

Table 2. Correlation between prediction accuracy (r_{MG}) and different combinations of factors that affect r_{MG} for different traits in 3217 population–trait combinations (BC_1F_2 populations excluded) in 969 maize biparental crosses.

Trait	Populations	$r^2(Nh^2)^{½}$	$E(r_{MG—r2})$	$(Nh^2)^{½}$	$E(r_{MG})$
Yield	799	0.30*b[†]B[‡]	0.32*aB	0.21*aA	0.21*bA
Moisture	898	0.45*cB	0.44*bB	0.36*bC	0.32*cdA
Test weight	850	0.43*cB	0.43*bB	0.34*bcA	0.34*dA
Stalk lodging	64	–0.02aA	–0.06cdA	–0.02adA	–0.06aeA
Root lodging	37	–0.04aA	–0.17dA	–0.20dA	–0.32eA
Plant height	358	0.23*abA	0.21*aA	0.22*aA	0.19*abA
Ear height	211	0.20*abA	0.19*acA	0.20*acA	0.18*abcA
All	3217	0.41	0.40	0.36	0.33

* Significant at the 0.05 probability level.

[†] Within columns (across different traits), correlation values followed by the same lowercase letter were not significantly different according to a Fisher z-transformations ($P = 0.05$; Bobko, 2001).

[‡] Within rows (across different factor combinations), correlation values followed by the same uppercase letter were not significantly different ($P = 0.05$). Significance tests were done by transforming the correlations to Fisher's z and considering that the correlations were nonindependent because they shared the same common variable r_{MG} (Bobko, 2001).

Table 3. Effective number of chromosome segments (M_e) back-calculated from Eq. [1] and [2] for seven different traits in 3034 population–trait combinations (r_{MG} between 0.1 and 1, BC_1F_2 populations excluded) from 969 maize biparental populations.

Trait	Populations	Mean N_M[†]	M_e from Li and Ji (2005)	Mean M_e from Eq. [1]	Mean M_e from Eq. [2]
Yield	742	70	59	468a[‡]	82c
Moisture	869	70	59	275b	29a
Test weight	811	70	59	283b	38ab
Stalk lodging	57	68	57	242ab	40ac
Root lodging	31	69	57	525ab	110ac
Plant height	338	72	60	434a	86c
Ear height	186	73	61	374ab	72bc

[†] N_M, number of markers used in genomewide prediction in each biparental population.
[‡] Within columns, values followed by the same letter were not significantly different according to a Tukey HSD test ($P = 0.05$).

Figure 1. Distribution of observed prediction accuracy (r_{MG}) in biparental maize populations for: (a) different traits for 3371 population–trait combinations; (b) different $r^2(Nh^2)^{1/2}$; (c) different expected r_{MG} from Eq. [2] [$E(r_{MG\_r2})$]; and (d) different $(Nh^2)^{1/2}$. Data in b, c, and d were for 3217 population–trait combinations for which BC_1F_2 populations were excluded. N = training population size, h^2 = heritability on an entry-mean basis, and $r^2 = r^2_{MM/2}$, where $r^2_{MM/2}$ was the mean linkage disequilibrium between a marker and QTL when the QTL was assumed to be at the midpoint of the two markers.

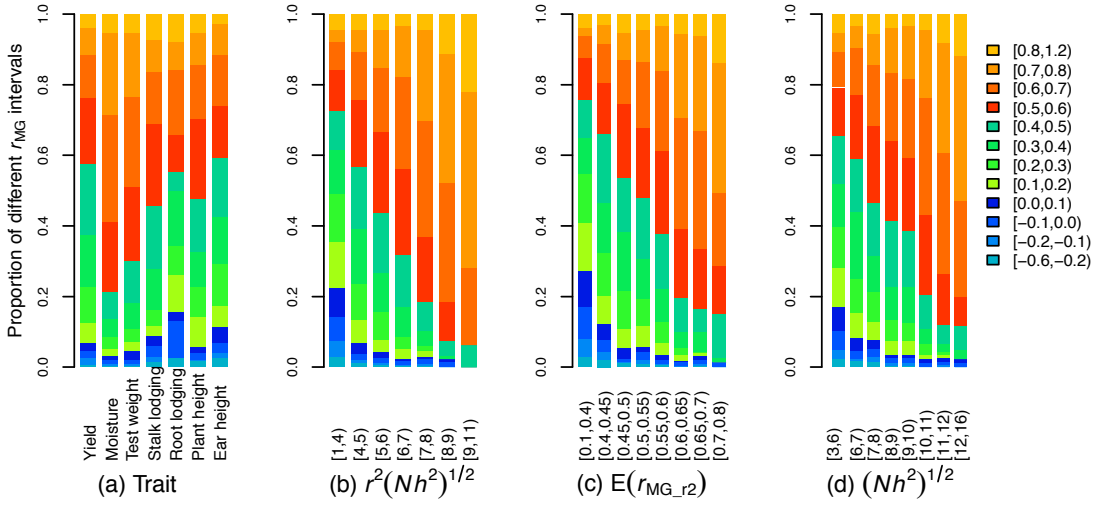


Figure 2. Observed prediction accuracy (r_{MG}) versus expected prediction accuracy from Eq. [1] [E(r_{MG})] and Eq. [2] [E(r_{MG_r2})] for 3217 population–trait combinations (BC_1F_2 populations excluded) in 969 maize biparental crosses.

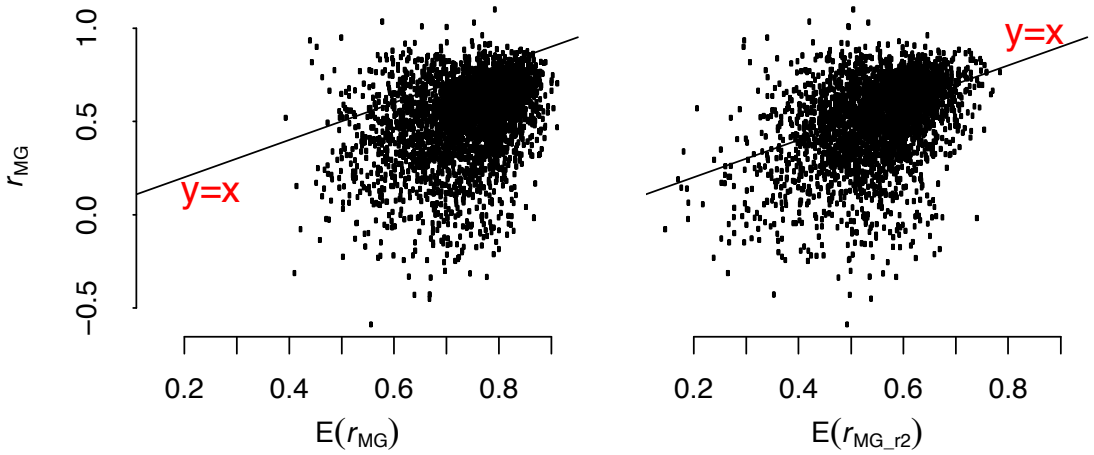
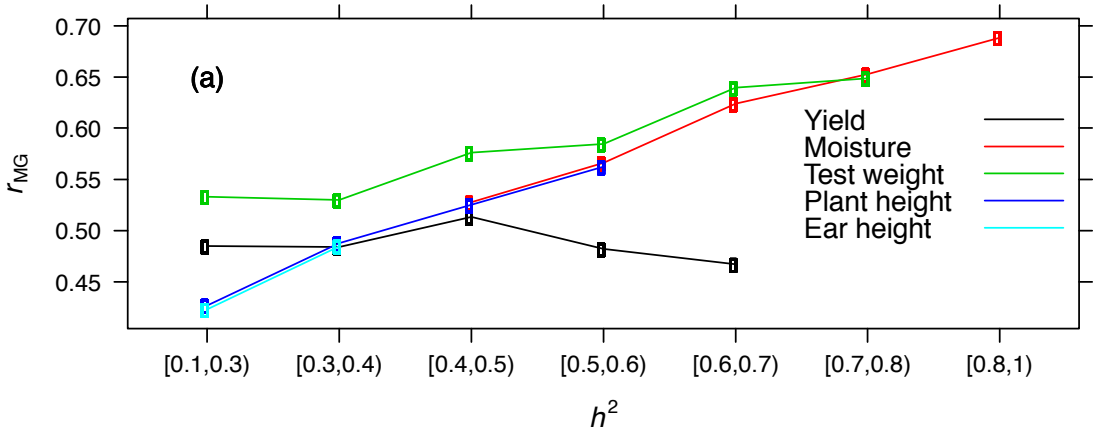
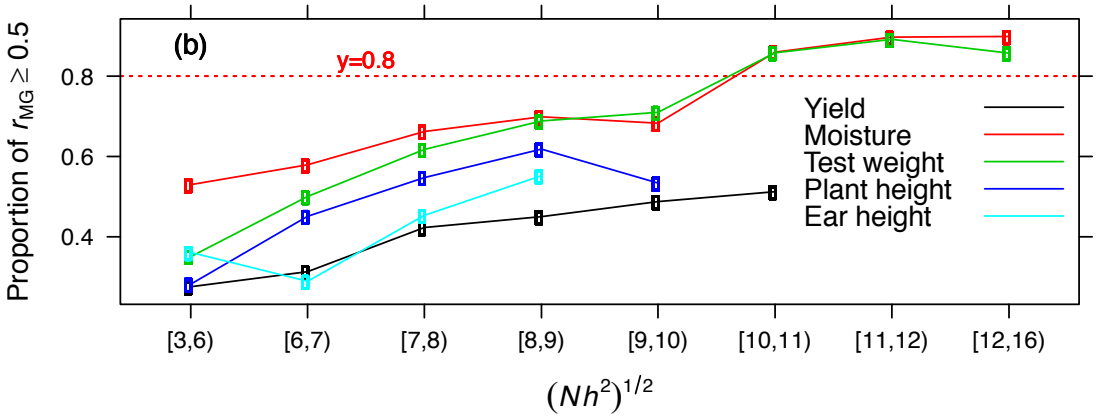
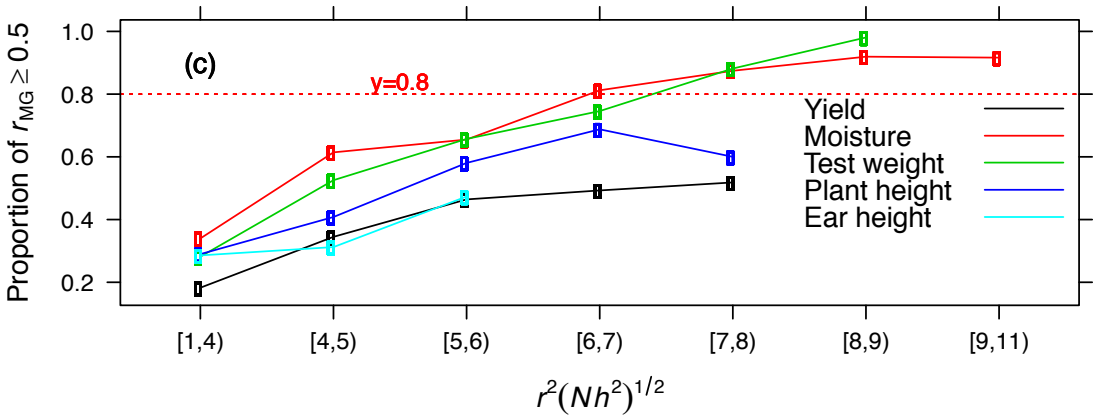


Figure 3. (a) Mean prediction accuracy (r_{MG}) in different intervals of entry-mean heritability (h^2) for five different traits in 1430 population–trait combinations with rounded training population sizes of $N_P = 180$; (b) proportion of r_{MG} equal to or larger than 0.50 in different intervals of $(Nh^2)^{1/2}$ for five different traits in 3062 population–trait combinations (BC_1F_2 populations excluded); and (c) proportion of r_{MG} equal to or larger than 0.50 in different intervals of $r^2(Nh^2)^{1/2}$ for five different traits in 3073 population–trait combinations (BC_1F_2 populations excluded). N = training population size, $r^2 = r^2_{MM/2}$, where $r^2_{MM/2}$ was the mean linkage disequilibrium between a marker and QTL when the QTL was assumed to be at the midpoint of the two markers. For all the above plots, population–trait combinations in an interval with < 20 data points were excluded.


(a)
0.70
0.65
0.60
0.55
0.50
0.45
r_{MG}
Yield
Moisture
Test weight
Plant height
Ear height
[0.1,0.3)
[0.3,0.4)
[0.4,0.5)
[0.5,0.6)
[0.6,0.7)
[0.7,0.8)
[0.8,1)
h^2


(b)
y=0.8
0.8
0.6
0.4
Proportion of $r_{MG} \geq 0.5$
Yield
Moisture
Test weight
Plant height
Ear height
[3,6)
[6,7)
[7,8)
[8,9)
[9,10)
[10,11)
[11,12)
[12,16)
$(Nh^2)^{1/2}$


(c)
y=0.8
1.0
0.8
0.6
0.4
0.2
Proportion of $r_{MG} \geq 0.5$
Yield
Moisture
Test weight
Plant height
Ear height
[1,4)
[4,5)
[5,6)
[6,7)
[7,8)
[8,9)
[9,11)
$r^2(Nh^2)^{1/2}$

Chapter 2: Prediction of genetic variance in biparental maize populations: Genomewide marker effects versus mean genetic variance in prior populations

Methods are lacking for predicting the genetic variance in biparental populations. Our objective was to determine whether genomewide marker effects and related populations could be used to predict the genetic variance (V_G) when two parents (A and B) are crossed to form a segregating population. For each of 85 A/B populations, 2–23 maize populations with A and B as one of the parents were used as the training population. In the genomewide selection model, the testcross V_G in A/B was predicted as the variance among the predicted genotypic values of progeny from the simulated A/B population. In the mean variance model, V_G in A/B was predicted as the mean of V_G in a series of A/* populations and B/* populations, where * denotes a random parent. The correlations between observed and predicted genetic variance were not significant ($P = 0.05$) for the genomewide selection model (–0.17 for yield, 0.12 for moisture and –0.06 for test weight) but were significant for the mean variance model (0.26 for yield, 0.46 for moisture, and 0.50 for test weight). The percentage of bias in estimates of V_G with the mean variance model was only 1% to 5%. Our results indicated that the V_G in an A/B population could be predicted as the mean variance among populations with A and B as one of the parents. The mean variance model should be practical in breeding programs because it simply utilizes phenotypic data from prior, related populations.

Introduction

In plant breeding, two elite parents typically are crossed and the best progeny are selected based on performance tests. The number of possible crosses can be very large. For example, 100 different parental inbreds lead to 4950 possible crosses. Because it is impossible to evaluate all of the possible crosses, it is desirable to predict the value of a cross before it is made.

The value of a biparental population is a function of the population mean as well as of the genetic variance (V_G) in the population (Schnell and Utz, 1975). In a self-pollinated species, the mean of the population can be predicted as the mean of the performance of the two parents. In a hybrid crop, for which the hybrid performance of the progeny when they are crossed with a common inbred (i.e., tester) is important, the testcross mean can be predicted as the mean of the testcross performance of the parents (Bernardo, 2010, p. 83). While methods are available for predicting the mean of progeny from a biparental population, good methods are lacking for predicting the V_G in a population. In this study, we considered two approaches for predicting the testcross V_G in maize (*Zea mays* L.) biparental populations: (1) a genomewide selection model [Zhong and Jannink (2007); Bernardo (2014)], in which genomewide markers are used to predict V_G; and (2) a mean variance model, in which V_G estimates in related populations are used to predict V_G.

Genomewide selection, which predicts the genotypic value of an individual from a large number of markers distributed across the genome (Meuwissen et al., 2001), has been found useful in plant breeding (Lorenzana and Bernardo, 2009; Lorenz et al., 2012; Poland et al., 2012; Zhao et al., 2012; Combs and Bernardo, 2013b; Riedelsheimer and

Melchinger, 2013). In genomewide selection, a test population is assumed to have been genotyped and a training population is assumed to have been genotyped and phenotyped. Marker effects estimated from the training population are used to predict the genotypic values of the test population for different quantitative traits. Suppose the test population comprises the progeny of the cross between inbreds A and B (denoted by A/B). A prior cross between parent A and any other inbred is denoted by A/*, and a prior cross between parent B and any other inbred is denoted by B/*. In a previous study, we found that genomewide selection in the A/B population is effective when the A/* and B/* populations are pooled into a training population for A/B (Jacobson et al., 2014).

In the genomewide selection model for predicting V_G, progeny from an A/B biparental population are simulated based on the sizes of the linkage groups, the map positions of the genomewide markers, and a given model for the incidence of crossing over during meiosis (Zhong and Jannink, 2007; Bernardo, 2014). The performance of each simulated progeny is predicted from the estimates of genomewide marker effects (from the A/* and B/* populations), and the V_G is then estimated as the variance among the marker-predicted genotypic values of the simulated progeny. This approach for predicting V_G therefore combines the use of simulated data (marker genotypes of simulated progeny in an A/B cross yet to be made) and nonsimulated data (prior estimates of marker effects from empirical experiments) to predict V_G. In contrast, in the mean variance model, the V_G in the A/B population is predicted as the mean V_G in the A/* and B/* populations.

The usefulness of the genomewide selection model and the mean variance model for predicting the V_G in an A/B population has not been investigated. Zhong and Jannink

(2007) and Bernardo (2014) proposed and demonstrated how genomewide markers can be used to predict V_G, but they did not have empirical data to test whether or not the estimates of V_G were accurate. In this study, we utilized phenotypic and marker data from 969 biparental maize populations in the Monsanto breeding program to investigate the genomewide selection model and mean variance model for predicting V_G. Because the data in this study represent the genetic backgrounds, traits, heritability (h^2), population sizes and extent of testing in a commercial maize breeding program, we expected our results to give a realistic picture of the extent to which V_G can be predicted. Our objective was to determine whether the V_G in an A/B population could be predicted by either genomewide selection model or mean variance model.

Materials and Methods

Overview of Models for Predicting Genetic Variance

In addition to the genomewide selection model and mean variance model for predicting V_G, we examined two other methods that provided a baseline for our comparisons. First, we estimated V_G by splitting the phenotypic data into two halves and determining the correspondence between the estimates of V_G from the two halves. This provided a measure of the repeatability of V_G estimated from field trials of progeny from the A/B population itself. Second, we used the molecular marker dissimilarity between parents A and B as a predictor of the relative amount of V_G among different A/B populations.

We also attempted to reduce the confounding effects of genotype-year interaction on the predictions of V_G. To do so, we predicted V_G with no restrictions on the years that

the field trials were conducted, as well as with the restriction that the A/B, A/*, and B/* populations were evaluated in the same year.

Phenotypic and Marker Data

Phenotypic and marker data for 969 biparental maize testcross populations were provided to us by Monsanto. A total of 485 A/B crosses were between inbreds from one heterotic group (Group 1) and 484 A/B crosses were between inbreds from an opposite heterotic group (Group 2). Among the 969 biparental populations, 707 were F_2 populations that were represented by F_3 families. For each F_2 population, plants grown from 10 to 15 selfed (F_3) seeds derived from each individual F_2 plant were testcrossed to an inbred tester from the opposite heterotic group. From the F_2 populations, we chose 85 A/B populations for comparing different methods of predicting V_G. All pedigrees in the dataset were coded by Monsanto to protect confidentiality.

The 85 A/B populations had 87–355 F_3 families. Testcrosses of the F_3 families were evaluated for grain yield (Mg ha^{-1}), moisture (g kg^{-1}), and test weight (kg hl^{-1}) at 4 to 12 locations in the U.S. from 2001 to 2008. All of the results (both means and variances) were therefore for testcross performance. Phenotypic data were available as the mean of each line within each location. Phenotypic data on some of the lines were missing from some locations, making the phenotypic data unbalanced.

Within each A/B population, the V_G and residual variance (V_R) were calculated for each trait by restricted maximum likelihood via the "lmer" function in the "lme4" package (Bates et al., 2013). The linear model was $y_{ij} = \mu + G_i + E_j + e_{ij}$, where y_{ij} was the testcross phenotypic value of the i_{th} line in the j_{th} environment, μ was the overall mean, G_i was the effect of i_{th} line, E_j was the effect of the j_{th} environment, and e_{ij} was the

residual effect. Because the data were entry means within each location, the genotype by environment interaction variance and within-location error variance were confounded in V_R. A likelihood ratio test was used to determine the significance of the estimates of V_G. The *p*-value from the likelihood ratio test was divided by 2.0 to approximate an F-test of the null hypothesis (Holland et al., 2003). The h^2 was estimated on an ad hoc basis as $h^2 = V_G/(V_G + V_R/l)$, where l was the mean number of locations (Holland et al., 2003). The value of l was estimated as the harmonic mean, given the unbalanced nature of the data (Holland et al., 2003).

The 85 A/B populations were chosen based on five criteria: (1) each population had at least one A/* population and one B/* population; (2) the A/B population was evaluated at four or more locations; (3) the A/B, A*, and B/* populations were F_2 populations that had at least 50 F_3 families and had a V_G significantly different from zero ($P = 0.05$); (4) the A/B, A/*, and B/* populations were crossed to the same tester; and (5) the accuracy of genomewide selection (r_{MP}, as described later) was significantly greater than zero ($P = 0.05$).

The parents of the A/B, A/*, and B/* populations were genotyped with 2911 single nucleotide polymorphism (SNP) markers, whereas the F_3 families within each population were genotyped with 38–103 markers. The SNP genotypes of an F_3 family were determined from a bulk of 10–15 F_3 plants. Progeny markers were imputed according to the conditional distribution of F_2 individuals for the 2911 SNP markers among the parents (Wu et al., 2007). Some marker genotypes were missing for the parents, and SNP marker genotypes were not imputed for these markers. Altogether, there were 2661–2893 markers for each population after imputation.

Repeatability of Estimates of Genetic Variance

For each A/B population, the testing locations were divided into two subsets, each subset with half of the locations. When the number of locations (l) was an odd number, the locations were divided into a subset of $(l - 1)/2$ locations and a subset of $(l + 1)/2$ locations. The V_G was calculated from subset 1 as V_{G1} and from subset 2 as V_{G2}. The correlation between V_{G1} and V_{G2} was calculated across the 85 A/B populations and was denoted by $r(V_{G1}, V_{G2})$.

To obtain $r(V_{G1}, V_{G2})$ among the 85 A/B populations, each A/B population needed estimates of V_{G1} and V_{G2} from a random combination. As previously mentioned, each A/B population was selected to have a minimum of four locations used in field trials. If an A/B population was tested in four locations, there were at most six unique combinations of $l/2 = 2$ locations for this population. In this case, even if some of the remaining 84 A/B populations were tested in more locations and therefore had more than six combinations (of half of the locations), there was no additional random combination from this A/B population that corresponded to those additional combinations from the other populations. Thus, $r(V_{G1}, V_{G2})$ could not be calculated for those additional combinations. Therefore, we only used six combinations of locations for each of the 85 A/B populations. For the populations with more than six combinations, a random sample of six combinations was selected from all the possible combinations. The $r(V_{G1}, V_{G2})$ was obtained by averaging the correlations across the six combinations.

Genomewide Selection Model

A total of 1000 F_2 plants were simulated for each A/B population; we did not need to simulate F_3 lines because the testcross performance of an F_2 plant is expected to

be equal to the testcross performance of its F_3 family (Bernardo, 2010, p. 217). The SNP genotypes of each A/B population were simulated based on the genotypes of parent A and parent B at the 2911 SNP loci. The sizes of each chromosome ranged from 1.18 Morgans to 2.57 Morgans. For each chromosome, a random haplotype was sampled as either from parent A or parent B, with crossing over occurring at random. The expected number of cross overs (L) in this haplotype was the length of the chromosome in Morgan, whereas the actual number of crossovers for the haplotype was sampled from a Poisson distribution with a mean of L. To account for interference, two adjacent crossovers were arbitrarily assumed to be at least 20 cM from each other. Our preliminary results (not shown) indicated that the frequency of crossovers affected the amount of V_G, but not the ranking of the A/B populations in terms of their V_G.

Genomewide marker effects were calculated according to a general combining ability model as described by Jacobson et al. (2014). For each A/B population, the corresponding A/* and B/* populations were pooled into the training population. The number of A/* and */B populations (N_X) for each of the 85 A/B test population ranged from 2 to 23 (Table 1). The total size of the pooled training population (N_T) ranged from 234 to 3736 individuals. The mean and range (in parentheses) of heritability for the 85 A/B populations were 0.43 (0.18, 0.73) for yield, 0.69 (0.36, 0.85) for moisture, and 0.59 (0.20, 0.87) for test weight.

Genomewide marker effects at the N_M marker loci were obtained by ridge regression-best linear unbiased prediction (Bernardo and Yu, 2007; Meuwissen et al., 2001). For a given trait, the testcross genotypic values of all N F_3 families in the A/B test population were predicted as $\hat{g} = \mu\mathbf{1} + \mathbf{Xm}$, where $\hat{g}$ was an $N \times 1$ vector of predicted

performance; μ was the estimated overall mean; **1** was an $N \times 1$ vector with elements equal to 1; **X** was an $N \times N_M$ matrix of simulated SNP-genotype indicators with elements of 1 if the SNP locus in the F_3 families was homozygous for the marker from parent A, –1 if the SNP locus was homozygous for the marker from parent B, and 0 if the SNP locus was heterozygous; and **m** was an $N_M \times 1$ vector of RR-BLUP marker effects averaged across the A/* and */B populations as described by (Jacobson et al., 2014). All the imputed markers in each A/* and */B population were used in RR-BLUP analysis within the population, and the N_M for obtaining $\hat{g}$ referred to the number of common markers in the simulated A/B population and imputed training populations. The genetic variance of each A/B population was predicted as $V_{\hat{g}} = \text{var}(\hat{g})$. The prediction accuracy for testcross genetic variance of A/B populations was calculated as the correlation $[r(V_G, V_{\hat{g}})]$ between $V_{\hat{g}}$ and V_G among all A/B test populations.

To evaluate the effectiveness of genomewide selection in predicting the performance of each F_3 family within each A/B population, we also calculated r_{MP}, which was the correlation between $\hat{g}$ and the phenotypic values of the F_3 families within each A/B population. In this case, the SNP genotypes of the actual F_3 families in the A/B population were used in obtaining the **X** matrix in the above linear model.

Mean Variance Model

The V_G estimates from each of the A/* populations were averaged to obtain $V_{G\bar{A}}$, and the V_G estimates from each of the B/* populations were averaged to obtain $V_{G\bar{B}}$. The genetic variance in each A/B population was then predicted as $V_{G\overline{AB}} = [V_{G\bar{A}} + V_{G\bar{B}}]/2$. The

prediction accuracy for genetic variance of A/B populations was calculated as the correlation [$r(V_G, V_{G\overline{AB}})$] between $V_{G\overline{AB}}$ and V_G among all A/B test populations.

Marker Dissimilarity as a Predictor of Relative Genetic Variance

The genetic similarity (S) between inbred parents A and B was calculated as the simple matching coefficient (Sneath and Sokal, 1973). The genetic dissimilarity between parents A and B was calculated as D = 1 – S. The correlation between V_G and D [$r(V_G, D)$] was calculated among all A/B populations.

Eliminating Genotype-Year Interaction

The training populations (A/* and B/*) and the A/B population may have been evaluated in different years. It is possible that the models for predicting V_G are inherently effective, but a substantial amount of genotype-year interaction would lead to reductions in the correlation between the predicted V_G (from one or more years) and the observed V_G (from a different year). Requiring all the A/* and B/* populations to be tested in the same year as the A/B population would remove any confounding effects of geneotype-year interaction on the accuracy of predicting V_G.

We therefore studied 37 A/B populations that allowed us to examine the influence of genotype-year interaction on the prediction of V_G. These 37 A/B populations had at least one A/* population and one B/* population tested in the same year as the A/B population. In the all-years analysis, both the genomewide selection model and the mean variance model were applied to the 37 A/B populations without any restrictions on the corresponding A/* and B/* populations. In the same-year analysis, the genomewide

selection model and the mean variance model were applied while considering only those A/* and B/* populations that were evaluated in the same year as the A/B population.

Across the 37 A/B populations, the number of A/* and B/* populations that were pooled into a training population (N_X) ranged from 2 to 23 in the all-years analysis and from 2 to 8 in the same-year analysis (Table 2). The total size of the pooled training population (N_T) ranged from 234 to 3763 F_3 families in the all-years analysis and from 234 to 1439 F_3 families in the same-year analysis (Table 2). The mean and range (in parentheses) of heritability for the 37 A/B populations were 0.43 (0.20, 0.64) for yield, 0.71 (0.51, 0.84) for moisture, 0.60 (0.26, 0.79) for test weight (Table 2).

Significance Tests

Significance tests for the accuracy of predicting V_G among the 85 or 37 A/B populations were performed by bootstrap sampling (Efron and Tibshirani, 1993). With the 85 A/B populations, for example, we took 1000 bootstrap samples, each of size 85 by sampling with replacement. For each of the 1000 bootstrap samples, we calculated $r(V_{G1}, V_{G2})$, $r(V_G, V_{G\overline{AB}})$, $r(V_G, V_{\hat{g}})$, $r(V_G, D)$ and their pairwise differences. A 95% bootstrap confidence interval was obtained by taking the 25th highest value (lower limit) and 975th highest value (upper limit) of the 1000 samples.

Results and Discussion

Baseline Predictions of V_G: Repeatability and Correlation with Marker Dissimilarity

The correlation between V_G estimated from one subset of environments (denoted by V_{G1}) and V_G estimated from another subset of environments (denoted by V_{G2}) was significant for moisture and test weight but not for yield (Table 3). Specifically, the $r(V_{G1},V_{G2})$ (with 95% confidence interval in parentheses) was 0.13 (–0.04, 0.29) for yield, 0.43 (0.28, 0.57) for moisture, and 0.55 (0.42, 0.66) for test weight. The $r(V_{G1}, V_{G2})$ was not significantly different between moisture and test weight.

The insignificant $r(V_{G1},V_{G2})$ for yield indicated the general intractability of predicting V_G for the most important trait in maize. Also, the insignificant $r(V_{G1},V_{G2})$ for yield was consistent with yield having the lowest h^2 among the three traits studied. If estimates of V_G for the same A/B population evaluated at different subsets of environments are largely nonrepeatable, it should be no surprise if predictions of V_G from related populations tested in different environments are likewise inaccurate.

Previous attempts to rank A/B populations for their V_G have relied on marker dissimilarity between the A and B parents, but the correlations between V_G and marker dissimilarity have been inconsistent (Moser and Lee, 1994; Moser and Lee, 1994; Manjarrez-Sandoval et al., 1997). In this study, the $r(V_{G,}D)$ was nonsignificant for yield and significant for moisture and test weight (Table 3). The observed $r(V_{G,}D)$ and 95% bootstrap confidence interval (in parenthesis) were 0.24 (–0.01, 0.48) for yield, 0.43 (0.28, 0.58) for moisture and 0.38 (0.17, 0.55) for test weight. Like previous studies, our results therefore indicated that molecular marker dissimilarity could provide useful

information on the relative V_G for some traits but not for others. Furthermore, there are two important limitations in using marker dissimilarity to predict V_G. First, the method can only predict the relative amount of V_G but not the actual V_G in different A/B populations. Second, the method assumes that the correlation between marker distance and V_G are the same across different traits. This means that a cross with the highest marker dissimilarity between the parents would have the highest relative predicted V_G for yield, moisture, test weight, and any other trait. In reality, a given cross may have the highest V_G for yield whereas a different cross may have the highest V_G for another trait.

Genomewide Selection and Mean Variance Models

The accuracy of genomewide selection for predicting the testcross performance of individual F_3 families was statistically significant for all three traits. The r_{MP} values had a mean and range (in parentheses) of 0.28 (0.14, 0.52) for yield, 0.48 (0.16, 0.75) for moisture, and 0.43 (0.13, 0.70) for test weight. In contrast, the correlation between the observed V_G and the V_G predicted through the genomewide selection model was not significant for any of the three traits. The $r(V_G, V_{\hat{g}})$ values (95% confidence intervals in parentheses) were –0.17 (–0.35, 0.04) for yield, 0.12 (–0.02, 0.28) for moisture, –0.06 (–0.22, 0.16) for test weight.

Whereas the genomewide selection model was ineffective in predicting V_G, the mean variance model lead to positive, significant correlations [denoted by $r(V_G, V_{G\overline{AB}})$] between observed V_G and predicted V_G for all three traits (Table 3). The $r(V_G, V_{G\overline{AB}})$ values (95% confidence intervals in parentheses) were 0.26 (0.07, 0.45) for yield, 0.46 (0.30, 0.60) for moisture, and 0.50 (0.36, 0.63) for test weight.

For all three traits, $r(V_{G,} V_{G\overline{AB}})$ was significantly different from $r(V_{G,} V_{\hat{g}})$ but was not significantly different from $r(V_{G1,} V_{G2})$. These results indicated that the mean variance model was more effective than the genomewide selection model, and that the V_G predicted by the mean variance model was comparable to the V_G predicted from field tests of the same A/B population in a different sample of environments. In addition, the genomewide selection model led to a large negative bias in the predicted V_G. The percent bias $\{[\text{mean}(\sqrt{V_{\hat{g}}})\text{-mean}(\sqrt{V_G})]/\text{mean}(\sqrt{V_G})\}$ in estimating $\sqrt{V_G}$ with the genomewide selection model was –82% for yield, –74% for moisture, and –76% for test weight. In contrast, the percent deviation between the predicted and observed $\sqrt{V_G}$ with the mean variance model was only 3% for yield, 5% for moisture, and 1% for test weight.

The reasons for the failure of the genomewide selection model are not entirely clear. Nevertheless, we propose two possible reasons for its ineffectiveness. First, the $V_{\hat{g}}$ includes a variance component due to error in estimating marker effects, which in turn leads to a prediction accuracy that is less than 1. The variance of this estimation error among A/B populations can be very large especially when adjacent markers are highly correlated. The RR-BLUP approach reduces the estimation error but introduces bias. The differences in marker genotypes between test and training population may magnify the influence of the estimation error and the bias. Second, the markers cannot account for the entirety of V_G, which is reflected by the low r_{MP} values. The variation in the r_{MP} values will contribute the variation in $V_{\hat{g}}$, which in turn reduces the correlation between V_G and $V_{\hat{g}}$ among the A/B populations

Elimination of Genotype-Year Interaction

Among the 37 A/B populations that were subjected to both all-years and same-year analysis, the accuracy of genomewide selection for predicting the testcross performance of individual F_3 families was statistically significant. In the all-years analysis, the mean and range (in parentheses) of r_{MP} was 0.31 (0.18, 0.52) for yield, 0.49 (0.21, 0.75) for moisture, 0.42 (0.13, 0.70) for test weight. In the same-year analysis, the mean and range of r_{MP} was 0.28 (0.16, 0.46) for yield, 0.44 (0.13, 0.62) for moisture, 0.38 (0.13, 0.62) for test weight.

Despite these statistically significant r_{MP} values, the correlations between the observed V_G and predicted V_G from the genomewide selection model were either negative or not significantly different from zero for both the all-years analysis and the same-year analysis (Table 4). For yield, the $r(V_G, V_{\hat{g}})$ was –0.43 with the all-years analysis and 0.04 with the same-year analysis; for moisture, the $r(V_G, V_{\hat{g}})$ was –0.12 for the all-years analysis and 0.04 for the same-year analysis; for test weight, the $r(V_G, V_{\hat{g}})$ was –0.12 for the all-years analysis and –0.10 for the same-year training populations. Restriction of the analysis within the same year, so that any genotype-year interaction is eliminated, therefore did not improve the effectiveness of the genomewide selection model for predicting V_G.

Neither did the elimination of genotype-year interaction improve the effectiveness of the mean variance model for predicting V_G (Table 4). For yield, the $r(V_G, V_{G\overline{AB}})$ was 0.59 with the all-years analysis and 0.46 with the same-year analysis; for moisture, the $r(V_G, V_{G\overline{AB}})$ was 0.29 with the all-years analysis and 0.35 with the same-year training

analysis; for test weight, the $r(V_G, V_{G\overline{AB}})$ was 0.45 with the all-years analysis and 0.37 with the same-year analysis. We were aware that the elimination of genotype-year interaction was confounded with a reduction in the size of the training population. The slightly lower values of $r(V_G, V_{G\overline{AB}})$ in the case of same-year training populations for yield might be due to the smaller number of training populations in same-year training populations. In addition, correlation coefficients estimated from small sample sizes (37 A/B populations instead of the original 85) have higher variances and may have led to a lower power to detect significant differences (Fisher and others, 1921).

Applications in Plant Breeding

The usefulness of a cross is determined by both its mean and genetic variance (Schnell and Utz, 1975). Previously attempts to rank A/B populations for their V_G have mainly relied on marker dissimilarity (Moser and Lee, 1994; Kisha et al., 1997; Manjarrez-Sandoval et al., 1997) or phenotypic differences between the parents (Busch et al., 1971; Moser and Lee, 1994; Utz et al., 2001). Phenotypic differences were not well correlated with the genetic variances while the correlation between marker dissimilarity and genetic variance was inconsistent across different traits or experiments. These inconsistencies may be partly due to the small number of A/B populations or limited number of lines within populations to get a good estimate of genetic variance: only five A/B crosses studied by Manjarrez-Sandoval et al., (1997), and only 28 lines in each A/B cross were used by Kisha et al., (1997) to estimate V_G in one experiment. Predicting V_G from parental contributions (Bernardo and Nyquist, 1998) and prior estimates of V_G, as well as from genomewide selection models (Zhong and Jannink, 2007; Bernardo, 2014) has been proposed but no empirical data were available to test the proposed methods.

In this study, we investigated the prediction of V_G in 85 A/B populations that were each represented by 87 to 355 F_3 families. The large sample size and the eliteness of the Monsanto breeding populations made the results informative and meaningful. We were aware that the A/B populations were not evaluated for the purpose of examining the ability to predict V_G: if that were the case, all A/B populations would have been evaluated in one large, balanced field trial. On the other hand, doing so is not realistic in breeding because phenotypic data from breeding programs as a whole are always from different environments. Hence, the data we used to predict V_G was representative of the situations encountered in real-life breeding programs.

Our results indicated that the mean variance model was the best for predicting V_G. While genomewide selection has been found useful for predicting the genotypic value of individuals (Lorenzana and Bernardo, 2009; Lorenz et al., 2012; Poland et al., 2012; Zhao et al., 2012; Combs and Bernardo, 2013b; Riedelsheimer and Melchinger, 2013), our results indicated that it is not useful for predicting V_G. Predicted V_G values tended to be more accurate with the mean variance model than with marker dissimilarity of parents, although the differences between the two methods were not statistically significant. One important advantage of using the mean variance model instead of marker dissimilarity is that marker dissimilarity can only predict the ranking of populations for V_G, whereas the mean variance model can provide an estimate of V_G.

The predicted V_G can help in selecting the best possible crosses via the superior progeny values (Zhong and Jannink, 2007), which is equal to $s = \mu + i_p\sqrt{V_G}$, where μ is the population mean, and i_p is the selection intensity with $p\%$ selected. The value of $\sqrt{V_G}$ in predicting the s also depends on the relative variance in μ and $\sqrt{V_G}$. Results

have suggested a much larger variance in μ than in $\sqrt{V_G}$, and that variation in s is therefore expected to be largely due to variation in the means of the crosses (Zhong and Jannink, 2007). The deviations between the observed and predicted V_G with the genomewide selection model (>90% of each trait) and the mean variance model (<5% for each trait) indicated that the mean variance model leads to more accurate predictions of V_G. Overall, our results indicate that for predicting V_G, the mean variance model would be useful as well as practical because it simply utilizes existing phenotypic data from prior, related crosses (A/* and B/*).

Table 1. Number and size of training populations, heritability (h^2), and correlation between marker predicted and test phenotypic values of F_3 families (r_{MP}) for the 85 A/B maize populations.

	Test populations											
			h^2			Training populations				r_{MP}		
	Locations	N[†]	Yield	Moisture	Test weight	**A/***	***/B**	N_x[‡]	N_T[§]	Yield	Moisture	Test weight
Mean	7	173	0.43	0.69	0.59	4	4	7	1188	0.28	0.48	0.43
Mininmum	4	87	0.18	0.36	0.20	1	1	2	234	0.14	0.16	0.13
Maximum	12	355	0.73	0.85	0.87	13	13	23	3736	0.52	0.75	0.70

[†] N, number of F_3 families in each of the A/B maize populations.

[‡] N_X, number of A/* and B/* in the training population.

[§] N_T, total number of F_3 families in the A/* and B/* training population.

Table 2. Number and size of training populations, heritability (h^2), and correlation between marker predicted and test phenotypic values of F_3 families (r_{MP}) for the 37 A/B maize populations that were subjected to both all-years and same-year analyses.

	Test populations					All-year analysis							Same-year analysis						
			h^2			Training populations				r_{MP}			Training populations				r_{MP}		
	Locations	$N^{\dagger}$	Yield	Moisture	Test weight	**A/***	***/B**	$N_x^{\ddagger}$	$N_T^{\S}$	Yield	Moisture	Test weight	**A/***	***/B**	$N_x^{\ddagger}$	$N_T^{\S}$	Yield	Moisture	Test weight
Mean	8	175	0.43	0.71	0.60	4	4	7	1269	0.31	0.49	0.42	2	2	4	622	0.28	0.44	0.38
Minimum	6	116	0.20	0.51	0.26	1	1	2	234	0.18	0.21	0.13	1	1	2	234	0.16	0.13	0.13
Maximum	12	355	0.64	0.84	0.79	13	13	23	3736	0.52	0.75	0.70	6	6	8	1439	0.46	0.62	0.62

† N, number of F_3 families in each of the A/B maize populations.

‡ N_X, number of A/* and B/* in the training population.

§ N_T, total number of F_3 families in the A/* and B/* training population.

Table 3. Accuracy of different methods for predicting accuracy of V_G (95% bootstrap confidence interval in parenthesis) among 85 A/B maize populations.

Methods	Correlation	Yield†	Moisture	Test weight
Repeatability of V_G	$r(V_{G1}, V_{G2})$	0.13 (–0.04, 0.29) bc†	0.43 (0.28, 0.57) def	0.55 (0.42, 0.66) f
Mean variance	$r(V_G, V_{G\overline{AB}})$	0.26 (0.07, 0.45) cd	0.46 (0.30, 0.60) def	0.50 (0.36, 0.63) ef
Genomewide selection	$r(V_G, V_{\hat{g}})$	–0.17 (–0.35, 0.04) a	0.12 (–0.02, 0.28) bc	–0.06 (–0.22, 0.16) ab
Marker dissimilarity	$r(V_G, D)$	0.24 (–0.01, 0.48) bcde	0.43 (0.28, 0.58) def	0.38 (0.17, 0.55) cdef

† Correlations with the same letter are not significant different from each other. The significance test was applied to all possible comparisons of correlations.

Table 4. Accuracy of the genomewide selection model [$r(V_G, V_{\hat{g}})$] and mean variance model [$r(V_G, V_{G\overline{AB}})$] for 37 A/B maize populations with the all-year and same-year analyses.

	Yield		Moisture		Test weight	
Analyses	$r(V_G, V_{\hat{g}})$	$r(V_G, V_{G\overline{AB}})$	$r(V_G, V_{\hat{g}})$	$r(V_G, V_{G\overline{AB}})$	$r(V_G, V_{\hat{g}})$	$r(V_G, V_{G\overline{AB}})$
All-year	–0.43 (–0.64, –0.20) † a‡	0.59 (0.37, 0.76) h	–0.12 (–0.32, 0.12) b	0.29 (–0.04, 0.60) cdefgh	–0.12 (–0.35, 0.18) bc	0.45 (0.26, 0.66) fgh
Same-year	0.04 (–0.37, 0.41) bcdf	0.46 (0.17, 0.70) eg	0.04 (–0.18, 0.41) bcde	0.35 (–0.05, 0.65) cdefgh	–0.10 (–0.48, 0.31) abc	0.37 (0.10, 0.62) defgh

† 95% bootstrap confidence interval in parenthesis

‡ Correlations with the same letter are not significant different from each other. The significance test was applied to all possible comparisons of correlations

Bibliography

Bates, D., M. Maechler, B. Bolker, and S. Walker. 2013. Lme4: Linear mixed-effects models using Eigen and S4. R package version 1.0-4. R Foundation for Statistical Computing, Vienna, Austria.

Bernardo, R. 1991. Retrospective index weights used in multiple trait selection in a maize breeding program. Crop Sci. 31:1174-1179. doi:10.2135/cropsci1991.0011183X003100050020x.

Bernardo, R. 2009. Genomewide selection for rapid introgression of exotic germplasm in maize. Crop Sci. 49:419-425. doi:10.2135/cropsci2008.08.0452.

Bernardo, R. 2010. Breeding for quantitative traits in plants. 2nd ed. Stemma Press, Woodbury.

Bernardo, R. 2014. Genomewide selection of parental inbreds: classes of loci and virtual biparental populations. Crop Sci. doi:10.2135/cropsci2014.01.0088.

Bernardo, R., and W. Nyquist. 1998. Additive and testcross genetic variances in crosses among recombinant inbreds. Theor. Appl. Genet. 97:116-121.

Bernardo, R., and J.M. Yu. 2007. Prospects for genomewide selection for quantitative traits in maize. Crop Sci. 47:1082-1090. doi:10.2135/cropsci2006.11.0690.

Bobko, P. 2001. Correlation and regression: applications for industrial organizational psychology and management. 2nd ed. Sage Publications, Inc., Thousand Oaks, CA.

Busch, R., K. Lucken, and R. Frohberg. 1971. F1 hybrids versus random F5 line performance and estimates of genetic effects in spring wheat. Crop Sci. 11:357-361.

Combs, E., and R. Bernardo. 2013a. Accuracy of genomewide selection for different traits with constant population size, heritability, and number of markers. The Plant Genome 6. doi:10.3835/plantgenome2012.11.0030.

Combs, E., and R. Bernardo. 2013b. Genomewide selection to introgress semidwarf corn germplasm into U.S. Corn Belt inbreds. Crop Sci. 53:1427-1435.

Crossa, J., P. Pérez, J. Hickey, J. Burgueño, L. Ornella, J. Cerón-Rojas, X. Zhang, S. Dreisigacker, R. Babu, Y. Li, D. Bonnett, and K. Mathews. 2013. Genomic prediction in CIMMYT maize and wheat breeding programs. Heredity. doi:10.1038/hdy.2013.16.

Daetwyler, H.D., B. Villanueva, and J.A. Woolliams. 2008. Accuracy of predicting the genetic risk of disease using a genome-wide approach. PLOS ONE 3:e3395. doi:10.1371/journal.pone.0003395.

Dekkers, J. 2007. Prediction of response to marker-assisted and genomic selection using selection index theory. J. Anim. Breed. Genet. 124:331-341. doi: 10.1111/j.1439-0388.2007.00701.x.

Goddard, M. 2009. Genomic selection: prediction of accuracy and maximisation of long term response. Genetica 136:245-257. doi: 10.1007/s10709-008-9308-0.

Efron, B., and R.J. Tibshirani. 1993. An introduction to the bootstrap. 1st ed. Chapman & Hall/CRC, Boca Raton.

Fisher, R.A. 1921. On the" probable error" of a coefficient of correlation deduced from a small sample. Metron. 1:3-32.

Hallauer, A. 1990. Methods used in developing maize inbreds. Maydica 35:1-16.

Hayes, B., P. Bowman, A. Chamberlain, and M. Goddard. 2009. Invited review: genomic selection in dairy cattle: progress and challenges. J. Dairy Sci. 92:433-443. doi: http://dx.doi.org/10.3168/jds.2008-1646.

Heffner, E.L., A.J. Lorenz, J.-L. Jannink, and M.E. Sorrells. 2010. Plant breeding with genomic selection: gain per unit time and cost. Crop Sci. 50:1681-1690. doi:10.2135/cropsci2009.11.0662.

Holland, J.B., W.E. Nyquist, and C.T. Cervantes-Martínez. 2003. Estimating and interpreting heritability for plant breeding: an update. Plant Breed. Rev. 22:9-112. doi:10.1002/9780470650202.ch2.

Jacobson, A., L. Lian, S. Zhong, and R. Bernardo. 2014. General combining ability model for genomewide selection in a biparental cross. Crop Sci. doi: 10.2135/cropsci2013.11.0774.

Kisha, T., C. Sneller, and B. Diers. 1997. Relationship between genetic distance among parents and genetic variance in populations of soybean. Crop Sci. 37:1317-1325.

Kohavi, R. 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. p. 1137-1145. *In*: International joint conference on artificial

intelligence, Montreal, Quebec, Canada. August 20-25, 1995. American Association for Artificial Intelligence, Montreal.

Li, J., and L. Ji. 2005. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. Heredity 95:221-227. doi:10.1038/sj.hdy.6800717.

Lorenz, A.J., K.P. Smith, and J.L. Jannink. 2012. Potential and optimization of genomic selection for Fusarium head blight resistance in six-row barley. Crop Sci. 52:1609-1621.

Lorenz, A.J. 2013. Resource allocation for maximizing prediction accuracy and genetic gain of genomic selection in plant breeding: a simulation experiment. G3: Genes| Genomes| Genetics 3:481-491. doi: 10.1534/g3.112.004911.

Lorenzana, R.E., and R. Bernardo. 2009. Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. Theor. Appl. Genet. 120:151-161. doi: 10.1007/s00122-009-1166-3.

Makowsky, R., N.M. Pajewski, Y.C. Klimentidis, A.I. Vazquez, C.W. Duarte, D.B. Allison, and G. de los Campos. 2011. Beyond missing heritability: prediction of complex traits. PLOS Genetics 7. doi: 10.1371/journal.pgen.1002051.

Manjarrez-Sandoval, P., T.E. Carter, D. Webb, and J. Burton. 1997. RFLP genetic similarity estimates and coefficient of parentage as genetic variance predictors for soybean yield. Crop Sci. 37:698-703.

Manolio, T.A., F.S. Collins, N.J. Cox, D.B. Goldstein, L.A. Hindorff, D.J. Hunter, M.I.

McCarthy, E.M. Ramos, L.R. Cardon, A. Chakravarti, J.H. Cho, A.E. Guttmacher, A. Kong, L. Kruglyak, E. Mardis, C.N. Rotimi, M. Slatkin, D. Valle, A.S. Whittemore, M. Boehnke, A.G. Clark, E.E. Eichler, G. Gibson, J.L. Haines, T.F. C. Mackay, S.A. McCarroll, and P.M. Visscher. 2009. Finding the missing heritability of complex diseases. Nature 461:747-753. doi:10.1038/nature08494.

Massman, J.M., H.-J.G. Jung, and R. Bernardo. 2013. Genomewide selection versus marker-assisted recurrent selection to improve grain yie

Meuwissen, T.H., B.J. Hayes, and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics. 157:1819-1829.

Moser, H., and M. Lee. 1994. RFLP variation and genealogical distance, multivariate distance, heterosis, and genetic variance in oats. Theor. Appl. Genet. 87:947-956.

Poland, J., J. Endelman, J. Dawson, J. Rutkoski, S. Wu, Y. Manes, S. Dreisigacker, J. Crossa, H. Sánchez-Villeda, M. Sorrells, and J.-L. Jannink. 2012. Genomic selection in wheat breeding using genotyping-by-sequencing. Plant Gen. 5:103-113.

R Development Core Team. 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.</bok>

Riedelsheimer, C., and A.E. Melchinger. 2013. Optimizing the allocation of resources for genomic selection in one breeding cycle. Theor. Appl. Genet. 126:2835-2848.

Senior, M., E. Chin, M. Lee, J. Smith, and C. Stuber. 1996. Simple sequence repeat markers developed from maize sequences found in the genbank database: Map

construction. Crop Sci. 36:1676–1683. doi:10.2135/cropsci1996.0011183X003600060043x

Scheet, P. and M. Stephens. 2006. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. Am. J. Hum. Genet. 78:629–644.

Schnell, F.W., and H.F. Utz. 1975. F1-leistung und elternwahl euphy-der züchtung von selbstbefruchtern. In: Bericht über die arbeitstagung der vereinigung österreichischer pflanzenzüchter. BAL Gumpenstein, Gumpenstein. p. 243-248.

Smith, J., T. Hussain, E. Jones, G. Graham, D. Podlich, S. Wall, and M. Williams. 2008. Use of doubled haploids in maize breeding: Implications for intellectual property protection and genetic diversity in hybrid crops. Mol. Breed. 22:51–59. doi:10.1007/s11032-007-9155-1

Sneath, P.H. A., and R.R. Sokal. 1973. Numerical taxonomy, the principles and practice of numerical classification. W. H. Freeman, San Francisco.

Utz, H., M. Bohn, and A. Melchinger. 2001. Predicting progeny means and variances of winter wheat crosses from phenotypic values of their parents. Crop Sci. 41:1470-1478.

Windhausen, V.S., G.N. Atlin, J.M. Hickey, J. Crossa, J.-L. Jannink, M.E. Sorrells, et al. 2012. Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments. G3: Genes| Genomes| Genetics 2:1427–1436. doi:10.1534/g3.112.003699

Wu, R., C.-X. Ma, and G. Casella. 2007. Statistical genetics of quantitative traits: linkage, maps, and QTL. Springer, New York.

Zhao, Y., M. Gowda, W. Liu, T. Würschum, H.P. Maurer, F.H. Longin, N. Ranc, and J.C. Reif. 2012. Accuracy of genomic selection in European maize elite breeding populations. Theor. Appl. Genet. 124:769-776.

Zhong, S., and J.-L. Jannink. 2007. Using quantitative trait loci results to discriminate among crosses on the basis of their progeny mean and variance. Genetics. 177:567-576.

Appendix

Appendix 1

Expected Prediction Accuracy [E(r_{MG_r2})] with Incomplete Linkage Disequilibrium ($r^2 < 1$) between a Marker and a Quantitative Trait Locus

Assume that N is the number of individuals (i = 1 to N) and N_M is the number of markers (j = 1 to N_M). For the jth marker-QTL pair, assume $\mathbf{m}_{.j}$ is a vector of genotypic indicators for the jth marker; $\mathbf{x}_{.j}$ is a vector of genotypic indicators for the jth QTL; r_j is the correlation between m_{ij} and x_{ij} for any i; b_j is the true QTL effect; $\hat{b}_j$ is the estimated QTL effect; μ is the overall mean; $\mathbf{y}$ is a vector of phenotypic values; and $\mathbf{e}$ is a vector of residual (error) effects. The $\mathbf{m}_{.j}$ vector is corrected for its mean so that $\bar{m}_{.j} = 0$. Each marker-QTL pair is assumed independent of the other pairs. Each QTL is assumed to have an equal variance and we also assume $\text{var}(x) = \text{var}(m)$. From linear regression,

$$\begin{aligned}\hat{b}_j &= (\mathbf{m}'_{.j}\mathbf{m}_{.j})^{-1}\mathbf{m}'_{.j}\mathbf{y} \\ &= (\mathbf{m}'_{.j}\mathbf{m}_{.j})^{-1}\mathbf{m}'_{.j}(\boldsymbol{\mu} + \mathbf{x}_{.j}b_j + \mathbf{e})\end{aligned}$$

Because we have $\bar{m}_{.j} = \sum_{i=1}^{N} m_{ij} = 0$, we will have $\mathbf{m}'_{.j}\boldsymbol{\mu} = 0$ and $\mathbf{m}'_{.j}\bar{\mathbf{x}}_{.j}b_j = 0$. We can then write $\hat{b}_j$ as:

$$\begin{aligned}\hat{b}_j &= (\mathbf{m}'_{.j}\mathbf{m}_{.j})^{-1}\mathbf{m}'_{.j}(\mathbf{x}_{.j}b_j - \bar{\mathbf{x}}_{.j}b_j + \mathbf{e}) \\ &= (\mathbf{m}'_{.j}\mathbf{m}_{.j})^{-1}\mathbf{m}'_{.j}((\mathbf{x}_{.j} - \bar{\mathbf{x}}_{.j})b_j + \mathbf{e}) \\ &= r_j b_j + (\mathbf{m}'_{.j}\mathbf{m}_{.j})^{-1}\mathbf{m}'_{.j}\mathbf{e}\end{aligned}$$

We assume the phenotypic variance is equal to 1 (Daetwyler et al., 2008) and we define the following:

$$g_i = \sum_{j=1}^{N_M} x_{ij} b_j$$

$$\text{var}(g_i) = \sum_{j=1}^{N_M} \text{var}(x) b_j^2 = h^2$$

$$\hat{g}_i = \sum_{j=1}^{N_M} m_{ij} \hat{b}_j$$

$$\mathbf{m}'_{.j} \mathbf{m}_{.j} = N \, \text{var}(m)$$

If we assume $r_j = r$ for all values of j, then

$$\begin{aligned}
&\text{cov}(g_i, \hat{g}_i) \\
&= \sum_{j=1}^{N_M} \text{cov}(x_{ij} b_j, m_{ij} \hat{b}_j) \\
&= \sum_{j=1}^{N_M} \text{cov}(x_{ij} b_j, m_{ij} r_j b_j + m_{ij} (\mathbf{m}'_{.j} \mathbf{m}_{.j})^{-1} \mathbf{m}'_{.j} \mathbf{e}) \\
&= \sum_{j=1}^{N_M} r_j^2 b_j^2 \, \text{var}(x) \\
&= \sum_{j=1}^{N_M} b_j^2 \, \text{var}(x) r^2 \\
&= h^2 r^2
\end{aligned}$$

We let $\hat{e}_j = (\mathbf{m}'_{.j} \mathbf{m}_{.j})^{-1} \mathbf{m}'_{.j} \mathbf{e}$, with $\text{var}(\hat{e}_j) = (N \, \text{var}(m))^{-1} \sigma_e^2$. With the assumption by Daetwyler et al. (2008) that $\sigma_e^2 = 1$,

$$
\begin{aligned}
\operatorname{var}(\hat{g}_i) &= \operatorname{var}(\sum_{j=1}^{N_M} m_{ij}\hat{b}_j) \\
&= \operatorname{var}[\sum_{j=1}^{N_M} m_{ij}(r_j b_j + \hat{e}_j)] \\
&= \operatorname{var}(\sum_{j=1}^{N_M} m_{ij} r_j b_j) + \operatorname{var}(\sum_{j=1}^{N_M} m_{ij}\hat{e}_j) \\
&= \sum_{j=1}^{N_M} \operatorname{var}(m_{ij} r_j b_j) + \sum_{j=1}^{N_M} \operatorname{var}(m_{ij}\hat{e}_j) \\
&= \sum_{j=1}^{N_M} r_j^2 b_j^2 \operatorname{var}(m) + \sum_{j=1}^{N_M} \operatorname{var}(m)\operatorname{var}(\hat{e}) \\
&= r^2 \sum_{j=1}^{N_M} b_j^2 \operatorname{var}(m) + \sum_{j=1}^{N_M} \operatorname{var}(m)(N \operatorname{var}(m))^{-1} \\
&= h^2 r^2 + N_M / N
\end{aligned}
$$

Finally,

$$
\begin{aligned}
\mathrm{E}(r_{\mathrm{MG\_r2}}) &= \operatorname{cov}(g_i, \hat{g}_i) / \sqrt{\operatorname{var}(g_i)\operatorname{var}(\hat{g}_i)} \\
&= r^2 h^2 / [h^2 (r^2 h^2 + N_M / N)]^{1/2} \\
&= r^2 [N h^2 / (r^2 N h^2 + N_M)]^{1/2}
\end{aligned}
$$

Appendix 2

Derivation of $r^2_{MM/2}$, the Squared Correlation between a Marker and QTL when the QTL Is Assumed at the Midpoint of Two Markers

If there are two markers, A and C, with a QTL (B) at the midpoint of the two markers, the recombination rate between AB and BC should be equal. Assuming no interference, the recombination rate between AC can be calculated as (Haldane, 1919): $c_{AC} = c_{AB} + c_{BC} - 2c_{AB}c_{BC} = 2c_{AB} - 2c_{AB}^2$. Therefore, $c_{AB} = [1-(1-2c_{AC})^{1/2}]/2$.

Consider the gametes produced by F_1 or F_2 plants. Assume x_1 is the allele at marker 1 and x_2 is the allele at marker 2 in a gamete; p_1 is the frequency of parent 1 allele at marker 1; and p_2 is the frequency of parent 1 allele at marker 2; q_1 is the frequency of the parent 2 allele at marker 1; and q_2 is the frequency of the parent 2 allele at marker 2. Alleles from parent 1 are coded as 1 and alleles from parent 2 are coded as 0. The r^2 is the squared correlation between marker haplotypes; c is the recombination fraction between two markers.

Given that $E(x_1) = p_1 = 0.5$, $E(x_2) = p_2 = 0.5$, and $E(x_1x_2)=c/2$:

$$r^2 = \text{cov}^2(x_1, x_2)/[\text{var}(x_1)\text{var}(x_2)]$$

$$= E^2[(x_1 - p_1)(x_2 - p_2)]/(p_1q_1p_2q_2)$$

$$= E^2[x_1x_2 - p_1x_2 - p_2x_1 + p_1p_2]/(p_1q_1p_2q_2)$$

$$= (c/2 - 0.25)^2/0.5^4$$

$= (2c - 1)^2$

For BC populations and DH populations derived from the F_1, the squared correlation between marker genotypes is the same as the squared correlation between haplotype genotypes of the gametes produced by F_1 plants. For DH populations derived from the F_2, the squared correlation between marker genotypes is the same as the squared correlation between haplotype genotypes of the gametes produced by F_2 plants.

For F_2 populations, assume: g_1 is the genotype at marker 1; g_2 is the genotype at marker 2; x_1 and x_2 are the haplotypes from one chromosome; x_1' and x_2' are the haplotypes from the homologous chromosome. For an F_2 population:

$$\begin{aligned}
\text{cov}(g_1, g_2) &= \text{cov}(x_1 + x_1', x_2 + x_2') \\
&= \text{cov}(x_1, x_2) + \text{cov}(x_1, x_2') + \text{cov}(x_1', x_2) + \text{cov}(x_1', x_2') \\
&= 2\text{cov}(x_1, x_2) \\
\text{var}(g_1) &= \text{var}(x_1 + x_1') \\
&= 2\text{var}(x_1) \\
\text{var}(g_2) &= \text{var}(x_2 + x_2') \\
&= 2\text{var}(x_2) \\
r^2(g_1, g_2) &= \text{cov}^2(g_1, g_2)/[\text{var}(g_1)\text{var}(g_2)] \\
&= \text{cov}^2(x_1, x_2)/[\text{var}(x_1)\text{var}(x_2)]
\end{aligned}$$

So, for an F_2 population, the squared correlation between marker genotypes is the same as the squared correlation between haplotypes of the gametes produced by F_1 plants. Such relationship does not apply to a BC_1F_2 population.

Based on the above analysis, if we know that the squared correlation between marker genotypes is r^2_{MM}, then for F_2, BC, DH populations from F_1 and DH populations derived from from the F_2, the following relationships exist: $r^2_{MM} = (1-2c_{MM})^2$ and $c_{MM/2} = [1-(1-2c_{MM})^{1/2}]/2 = [1-(r^2_{MM})^{1/4}]/2$. We can therefore calculate the correlation squared between the marker and the midpoint QTL as $r^2_{MM/2} = (1-2c_{MM/2})^2 = [(r^2_{MM})^{1/4}]^2 = \sqrt{r^2_{MM}}$.

References:

Daetwyler, H.D., B. Villanueva, and J.A. Woolliams. 2008. Accuracy of predicting the genetic risk of disease using a genome-wide approach. PLOS One. 3:e3395.

Haldane, J. 1919. The combination of linkage values and the calculation of distances between the loci of linked factors. J. Genet. 8:309-320.